



VU Research Portal

A consistent conditional moment test of functional form

Bierens, H.J.

1989

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Bierens, H. J. (1989). *A consistent conditional moment test of functional form*. (Serie Research Memoranda; No. 1989-64). Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

ET

05348

1989
64

SERIE RESEARCH MEMORANDA

A CONSISTENT CONDITIONAL MOMENT TEST OF FUNCTIONAL FORM

by

Herman J. Bierens

Research Memorandum 1989-64

September 1989



**VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
EN ECONOMETRIE
AMSTERDAM**

A CONSISTENT CONDITIONAL MOMENT TEST OF FUNCTIONAL FORM

by Herman J. BIERENS ¹⁾

1. INTRODUCTION

Conditional moment (CM) tests of functional form exploit the property that for correctly specified models the conditional expectation of certain functions of the observations should be almost surely equal to zero. A chi-square misspecification test can then be based on weighted means of these functions. As has been shown by Newey (1985), most model misspecification tests are special forms of CM tests.

The power of the CM test depends heavily on the choice of the weighting functions. In particular, the CM test is not consistent against all possible alternatives. Since the CM test imposes only finitely many moment conditions, it is always possible to construct alternative data generating processes for which these moment conditions hold while the null is false.

To the best of our knowledge the only consistent model misspecification tests are those of Bierens (1982,1984,1987,1988) and Bierens and Hartog (1988). The tests of Bierens (1982,1984) are genuine consistent tests, but the null distribution of the test statistics involved is intractable and had to be approximated using Chebishev's inequality for first moments. The tests of Bierens (1987, 1988) and Bierens and Hartog (1988) have tractable null distributions, but their consistency is due to randomization of test parameters.

In the present paper it will be shown that any CM test of functional form of nonlinear regression models can be converted into a chi-square test that is consistent against *all* deviations from the null. The consistency of this test does not rely on randomization.

The plan of the paper is as follows. In Section 2 we state the hypotheses to be tested. In Sections 3 and 4 we show how to convert the CM test into a consistent test. In Section 5 we present and interpret the results of a limited Monte Carlo analysis. Finally, in Section 6 we show what kind of information about the true model the test provides if the null hypothesis is rejected. Appendix A contains formal statements of the assumptions maintained in our analysis. These assumptions are jointly referred to as "Assumption A". Appendix B contains the proofs of the lemmas.

2. THE HYPOTHESES TO BE TESTED

In developing our consistent version of the CM test we confine our attention to a random sample $\{(y_1, x_1), \dots, (y_n, x_n)\}$ from a distribution $F(y, x)$ on $R \times R^k$ for which $\int y^2 dF(y, x) < \infty$. It seems possible to extend the results below to the heterogeneous and/or time series case, using the approach in Bierens (1981, Ch.3, 1984, 1987, 1988), but incorporating these extensions in the present paper would diverge attention from the main theme.

In parametric regression analysis it is assumed that the regression function $g(x) = E(y_j | x_j = x)$ belongs to a parametric family of known real functions $f(x, \theta)$ on $R^k \times \Theta$, where $\Theta \subset R^m$ is the parameter space. Denote by $D(g)$ the set of all probability distribution functions $F(y, x)$ on $R \times R^k$ such that for a random drawing (y, x) from F , $E[y^2] < \infty$ and $P[E(y|x) = g(x)] = 1$. The null hypothesis to be tested is that the parametric specification involved is correct:

- (1) H_0 : The distribution F belongs to the set $D_0 = \bigcup_{\theta \in \Theta} D(f(\cdot, \theta))$.

In other words, the data generating process characterized by F is such that $P[E(y_j | x_j) = f(x_j, \theta_0)] = 1$ for some $\theta_0 \in \Theta$. The alternative hypothesis we wish to test is that the null is false, i.e.,

- (2) H_1 : The distribution F belongs to the set $D_1 = (\bigcup_g D(g)) \setminus D_0$,

where the union is over all Borel measurable real functions g on R^k . This is equivalent to the statement that F belongs to the class of distributions for which $P[E(y_j | x_j) = f(x_j, \theta)] < 1$ for all $\theta \in \Theta$.

Given a significance level α and corresponding critical region C_α , the asymptotic power function $\rho(F) = \lim_{n \rightarrow \infty} P(\hat{W} \in C_\alpha)$ of a test of H_0 with test statistic \hat{W} depends on the distribution F . Clearly, $F \in D_0$ implies $\rho(F) = \alpha$. If $\rho(F) = 1$ for all $F \in D_1$, then the test involved is said to be consistent.



3. TOWARDS A CONSISTENT CONDITIONAL MOMENT TEST

Newey's (1985) CM test of functional form of the nonlinear regression model under review imposes a finite number (say p) of moment conditions of the form

$$(3) \quad E[(y_1 - f(x_1, \theta_0))w_i(x_1, \theta_0)] = 0, \quad i=1, 2, \dots, p,$$

where

$$(4) \quad \theta_0 = \operatorname{argmin}_{\theta \in \Theta} E([y_1 - f(x_1, \theta)]^2)$$

and the $w_i(x, \theta)$ are weighting functions. Clearly, under H_0 the moment conditions (3) always hold. The weighting functions $w_i(x, \theta)$ are chosen such that under H_1 at least one of the conditions (3) (likely) fails to hold. Obviously, the more weighting functions $w_i(x, \theta)$ we use, the more likely these moment conditions will be violated under H_1 .

Under mild regularity conditions [cf. Jennrich (1969), White (1981, 1982) and appendix A] the nonlinear least squares estimator $\hat{\theta}$ is a consistent and asymptotically normally distributed estimator of θ_0 , even if the model is misspecified. A chi-square test can now be based on the sample moments $(1/n) \sum_{j=1}^n (y_j - f(x_j, \hat{\theta}))w_i(x_j, \hat{\theta})$, $i=1, 2, \dots, p$.

Most model specification tests of functional form can be put in this framework. For example, Ramsey's (1969, 1970) model specification tests are special cases of the CM test, and so is White's (1981) version of Hausman's (1978) test. See also Ruud (1984) for a review of Hausman-type tests and Newey (1985) for other examples of CM tests.

As mentioned before, the power of the CM test depends heavily on the choice of the weighting functions. In particular, this test cannot be consistent against all possible alternatives, due to the fact that only finitely many moment conditions are imposed. This suggests the use of an infinite set of moment conditions as a possible solution of the inconsistency problem. The following fundamental lemma indicates what kind of moment conditions are suitable.

LEMMA 1: Let v be a random variable or vector satisfying $E|v| < \infty$ ²⁾ and let x be a bounded random vector in R^k such that $P[E(v|x)=0] < 1$. Then the set $S = \{t \in R^k: E[v \cdot \exp(t'x)] = 0\}$ has Lebesgue measure zero.

Assume that the model is misspecified. Then $P\{E[y_1 - f(x_1, \theta_0) | x_1] = 0\} < 1$. Let Φ be an arbitrary Borel measurable bounded one-to-one mapping from R^k into R^k . For example, we may choose

$$(5) \quad \Phi(x) = \Phi(x^{(1)}, \dots, x^{(k)}) = (\tan^{-1}(x^{(1)}), \dots, \tan^{-1}(x^{(k)}))'.$$

Then conditioning on x_1 is equivalent to conditioning on the bounded random vector $\Phi(x_1)$, for x_1 and $\Phi(x_1)$ generate the same Borel field. Thus:

$$(6) \quad P\{E[y_1 - f(x_1, \theta_0) | \Phi(x_1)] = 0\} < 1.$$

It follows now from (6) and Lemma 1 that the set

$$(7) \quad S = \{t \in R^k: E[(y_1 - f(x_1, \theta_0)) \exp(t' \Phi(x_1))] = 0\}$$

has Lebesgue measure zero. This suggests to use $\exp[t' \Phi(x)]$ in place of the $w_1(x, \theta)$, i.e., one may base a consistent CM test on the single sample moment

$$(8) \quad \hat{M}(t) = (1/n) \sum_{j=1}^n (y_j - f(x_j, \hat{\theta})) \exp(t' \Phi(x_j)),$$

for under H_1 , $\text{plim}_{n \rightarrow \infty} \hat{M}(t) \neq 0$ for all t except in a set with Lebesgue measure zero.

In the sequel we shall derive a consistent CM test based on (8) only. A more general consistent CM test can be based on the sample moments

$$(9) \quad \hat{M}_i(t) = (1/n) \sum_{j=1}^n (y_j - f(x_j, \hat{\theta})) w_i(x_j, \hat{\theta}) \exp(t' \Phi(x_j)), \quad (i=1, \dots, p).$$

Despite the fact that the weighting functions w_i are irrelevant for consistency, it may make sense to consider this case as well. The weighting functions w_i determine a class of (implicit) alternative hypotheses against which the CM test has maximal power. Cf. Holly (1982). If these alternative

hypotheses are of special interest, one may wish to direct the power of the consistent test towards these alternatives. Moreover, since consistency is only an asymptotic property, the small sample power may be enhanced by using these weighting functions. By mimicking the logic of this paper, it is quite easy to derive these more general consistent CM tests.

Observe that if the model contains a constant term then $0 \in S$, for by the first-order condition for (4), $E(y_1 - f(x_1, \theta_0)) = 0$. Moreover, since S is just the set of contours at zero level of a continuous function, it cannot be a dense subset of R^k . In other words, for every $t_0 \notin S$ there exists an open neighborhood of t_0 with no points in S . Summarizing:

THEOREM 1: *Let Assumption A hold. Under H_1 the set S defined in (7) has Lebesgue measure zero and is not dense in R^k .*

Let us assume that H_0 is true. Denote

$$(10) \quad b(t) = E[(\partial/\partial\theta')f(x_1, \theta_0)\exp(t'\Phi(x_1))];$$

$$(11) \quad A = E[(\partial/\partial\theta')f(x_1, \theta_0)](\partial/\partial\theta)f(x_1, \theta_0).$$

It is a standard exercise in asymptotic theory to verify [cf. the proof of Lemma 3 below] that under H_0 and Assumption A,

$$(12) \quad \sqrt{n}M(t) \rightarrow N[0, s^2(t)] \text{ in distribution,}$$

pointwise in t , where

$$(13) \quad s^2(t) = E\{(y_1 - f(x_1, \theta_0))^2 [\exp(t'\Phi(x_1)) - b(t)'A^{-1}(\partial/\partial\theta')f(x_1, \theta_0)]^2\}.$$

Note that $s^2(0) = 0$ if the model contains a constant term. The function $s^2(t)$ can be consistently estimated by

$$(14) \quad \hat{s}^2(t) = (1/n)\sum_{j=1}^n (y_j - f(x_j, \hat{\theta}))^2 [\exp(t'\Phi(x_j)) - \hat{b}(t)'\hat{A}^{-1}(\partial/\partial\theta')f(x_j, \hat{\theta})]^2,$$

where

$$(15) \hat{b}(t) = (1/n) \sum_{j=1}^n [(\partial/\partial \theta') f(x_j, \hat{\theta}) \exp(t' \Phi(x_j))],$$

$$(16) \hat{A} = (1/n) \sum_{j=1}^n ((\partial/\partial \theta') f(x_j, \hat{\theta})) ((\partial/\partial \theta) f(x_j, \hat{\theta})).$$

Now let

$$(17) \sigma^2(x_j) = E[(y_j - f(x_j, \theta_0))^2 | x_j]$$

and assume:

ASSUMPTION B: $P[\sigma^2(x_1) > 0] = 1$. There exists a Borel measurable real function μ on R^k such that the random vector $\kappa = (\mu(x_1), (\partial/\partial \theta) f(x_1, \theta_0))'$ has non-singular second moment matrix $E[\kappa \kappa']$.

Then

LEMMA 2: Under Assumption B the set $S_* = \{t \in R^k : s^2(t) = 0\}$ has Lebesgue measure zero and is not dense in R^k .

This result holds under H_1 as well. The importance of Lemma 2 is that the statistic

$$(18) \hat{W}(t) = n[\hat{M}(t)]^2 / \hat{s}^2(t)$$

is well-defined, possibly except for t in the set SUS_* with Lebesgue measure zero.

From Theorem 1 and Lemma 2 it now easily follows:

THEOREM 2: Let Assumptions A-B hold. There exists a non-dense subset S of R^k with Lebesgue measure zero such that for every $t \in R^k \setminus S$, $\hat{W}(t) \rightarrow \chi_1^2$ in distribution under H_0 , whereas under H_1 , $\hat{W}(t)/n \rightarrow \eta(t)$ a.s., where $\eta(t) > 0$.

It should be noted that the set S in Theorem 2 depends on the distribution F of (y_j, x_j) . This implies that in general we cannot choose a fixed t for which the test is consistent. Nevertheless, the result of Theorem 2

is close to a genuine consistent test, as will be shown in Section 4.

REMARK ON ASSUMPTION B: The first part of Assumption B is hardly a condition. The second part actually states that there exists a function $\mu(x)$ such that the parameters θ and α of the augmented model $y_j = f(x_j, \theta) + \alpha \cdot \mu(x_j) + u_j$ are locally identifiable. If the model does not contain a constant term a possible choice for μ may be $\mu(x)=1$. In that case Assumption B is simply the local identification condition that is typically assumed in nonlinear least squares estimation for models with a constant term. Of course, if the model already contains a constant term the choice $\mu(x)=1$ will not work. The existence of a suitable function μ depends on the variation in x_1 . Consider for example the case where x_1 takes only as many values as the dimension m of θ_0 . Then any function of x_1 can be written as a linear combination of $(\partial/\partial\theta)f(x_1, \theta_0)$, so that in this case it is impossible to find a μ for which Assumption B holds.

REMARK ON LEMMA 1: Lemma 1 yields as byproduct the following general series expansion of conditional expectation functions:

COROLLARY 1: Let v be a random variable satisfying $E[v^2] < \infty$ and let x be a random vector in R^k . For any Borel measurable bounded one-to-one mapping Φ from R^k into R^k and any sequence (t_j) , $j=1,2,\dots$, in R^k that is dense in a set $T \subset R^k$ with positive Lebesgue measure there exist coefficients $\beta_{n,j}$, $j=0,\dots,n$, $n=0,1,2,\dots$, such that

$$E(v|x) = \beta_{0,0} + \sum_{n=1}^{\infty} [\beta_{n,0} + \sum_{j=1}^n \beta_{n,j} \exp(t_j' \Phi(x))] \text{ a.s.}$$

PROOF: Without loss of generality we may assume that x is bounded itself, so that we may choose $\Phi(x)=x$. For $n=1,2,\dots$, let

$$f_n(x) = \alpha_{n,0} + \sum_{j=1}^n \alpha_{n,j} \exp(t_j' x)$$

where $\alpha_{n,n} = 1$ and the other $\alpha_{n,j}$'s are chosen such that

$$E f_n(x) = 0, \quad E[f_n(x) \exp(t_j' x)] = 0 \text{ if } j < n.$$

This is always possible. Now define the function $\psi_n(z)$ on the range Z of x by $\psi_n(z) = 1$ for $n = 0$ and

$$\psi_n(z) = f_n(z)/[E(f_n(x)^2)]^{1/2} \text{ if } E[f_n(x)^2] > 0, \psi_n(z) = 0 \text{ if } E[f_n(x)^2] = 0$$

for $n > 0$, and let $\gamma_n = E[v \cdot \psi_n(x)]$. The nonzero functions ψ_n form an orthonormal system of the Hilbert space H of Borel measurable functions φ on Z satisfying $E[\varphi(x)^2] < \infty$, with inner product $(\psi, \varphi) = E[\psi(x)\varphi(x)]$. Moreover, the coefficients γ_n are the Fourier coefficients of the function $h \in H$ defined by $h(x) = E(v|x)$. According to Royden (1968, p.212) there exists an element g of H such that $g(z) = \sum_{n=0}^{\infty} \gamma_n \psi_n(z)$, which can be rewritten as

$$g(z) = \beta_{0,0} + \sum_{n=1}^{\infty} [\beta_{n,0} + \sum_{j=1}^n \beta_{n,j} \exp(t_j'z)].$$

Since $E(v-g(x))\exp(t_j'x) = 0$ for $j=1,2,\dots$, $E(v-g(x))\exp(t'x)$ is continuous in t and (t_1, t_2, \dots) is dense in T , we have $E(v-g(x))\exp(t'x) = 0$ for all $t \in T$. Since T has positive Lebesgue measure, it follows from Lemma 1 that $g(x) = E(v|x)$ a.s. Q.E.D.

Note that if v is interpreted as a regression residual the test in Theorem 2 (and Theorems 3, 4 and 5 below) actually tests the null hypothesis that all the $\beta_{n,j}$'s are zero. Finally, note that this result is reminiscent of the Fourier expansion approach of Gallant. See Gallant (1981, 1982, 1984) and El Badawi, Gallant and Souza (1983). Also, it is related to projection pursuit regression. See Friedman and Stuetzle (1981) and Huber (1985).

4. THE CHOICE OF t AND Φ

The choice of t

Since the power of the test in Theorem 2 depends heavily on t , we might think of maximizing $\hat{W}(t)$ over some subset T of R^k and using the resulting \hat{t} instead of t . However, the resulting test statistic is not necessarily asymptotically χ_1^2 distributed under H_0 . On the other hand, it will be shown that if T is a hypercube in R^k then $\sup_{t \in T} \hat{W}(t)$ does converge in distribution under H_0 , but its limiting distribution depends on the joint distribution of (y_j, x_j) . The latter implies that it is not possible to calculate critical values of the test $\sup_{t \in T} \hat{W}(t)$ that are generally applicable. This problem can be overcome by choosing between \hat{t} and a fixed t , where a penalty function is introduced on choosing \hat{t} . Under H_0 this penalty is effective, by which the test becomes asymptotically equivalent to $\hat{W}(t)$ with fixed t . Under H_1 the penalty is ineffective, leading asymptotically to the choice of $\sup_{t \in T} \hat{W}(t)$ as test statistic.

In the sequel we shall choose T such that

$$T = \times_{i=1}^k [\tau_{1i}, \tau_{2i}], \text{ with } -\infty < \tau_{1i} < \tau_{2i} < \infty \text{ and } s^2(t) > 0 \text{ for } t \in T.$$

Although Assumption B guarantees that the set $\{t \in R^k : s^2(t) > 0\}$ contains a compact subset T with positive Lebesgue measure, in practice we can only choose T freely if we assume that $s^2(t) > 0$ for all t (except $t=0$ if the model contains a constant term). From the proof of Lemma 2 it is clear that this is very weak a condition. In generalizing Theorem 2 and the theorems below to other CM tests we have to make a similar assumption, namely that the asymptotic variance matrix of the vector of sample moments (9) times \sqrt{n} is nonsingular for all t (or all $t \neq 0$).

Before we proceed, let us first briefly review some terminology related to convergence of probability measures on metric spaces. For a full account, see Royden (1968) and Billingsley (1968). Let $C(T)$ be the metric space of all continuous real functions on T , with metric $\rho(z_1, z_2) = \sup_{t \in T} |z_1(t) - z_2(t)|$. The Borel sets of $C(T)$ are the members of the σ -Algebra generated by the open sets in $C(T)$. Since $\hat{W}(t)$ is a.s. continuous, it is a stochastic element of $C(T)$, and so is $\sup_{t \in T} \hat{W}(t)$. Let (z_n) be a sequence of stochastic elements of $C(T)$. Each z_n induces a probability measure P_n on $C(T)$ by the correspondence $P_n(B) = P\{z_n \in B\}$, where B is an

arbitrary Borel set in $C(T)$. We say that P_n converges weakly to P if $\lim_{n \rightarrow \infty} P_n(B) = P(B)$ for each Borel set B in $C(T)$ with boundary ∂B satisfying $P(\partial B) = 0$. If P is the probability measure induced by a stochastic element z of $C(T)$ then we also say that z_n converges weakly to z . A necessary condition for weak convergence is that P_n is tight, i.e., for every $\epsilon \in (0,1)$ there exists a compact subset K of $C(T)$ such that $\sup_n P_n(K) > 1-\epsilon$. We say that z_n is tight if P_n is tight. A stochastic element z of $C(T)$ is Gaussian with covariance function $\Gamma(t_1, t_2)$ if for arbitrary q and t_1, \dots, t_q in T , $(z(t_1), \dots, z(t_q))'$ is q -variate normally distributed with zero mean vector and variance matrix $(\Gamma(t_i, t_j))$, $i, j=1, \dots, q$. Gaussian elements of $C(T)$ are fully characterized by their covariance functions.

THEOREM 3: *Let Assumptions A-B hold and let H_0 be true. Then \hat{W} converges weakly to z^2 , where z is a Gaussian element of $C(T)$ with covariance function*

$$\begin{aligned} \Gamma(t_1, t_2) = & E\{(y_1 - f(x_1, \theta_0))^2 [\exp(t_1' \Phi(x_1)) - b(t_1)' A^{-1}(\partial/\partial \theta') f(x_1, \theta_0)] \\ & \times [\exp(t_2' \Phi(x_1)) - b(t_2)' A^{-1}(\partial/\partial \theta') f(x_1, \theta_0)]\} / ((\sqrt{s^2(t_1)})(\sqrt{s^2(t_2)})). \end{aligned}$$

Moreover, $\hat{W}(t)$ with $\hat{t} = \arg\max_{t \in T} \hat{W}(t)$ converges in distribution to $\sup_{t \in T} z(t)^2$. ³⁾ Furthermore, under Assumptions A-B and H_1 , $\hat{W}(t)/n \rightarrow \eta(t)$ a.s. uniformly on T and consequently $\sup_{t \in T} \hat{W}(t)/n \rightarrow \sup_{t \in T} \eta(t)$ a.s., where η is defined in Theorem 2.

PROOF: The result under H_1 follows straightforwardly from the uniform law of large numbers of Jennrich (1969). The result under H_0 is based on the following two lemmas. Let $u_j = y_j - f(x_j, \theta_0)$ and let

$$z_n(t) = (1/\sqrt{n}) \sum_{j=1}^n u_j [\exp(t' \Phi(x_j)) - b(t)' A^{-1}(\partial/\partial \theta') f(x_j, \theta_0)]^2 / s^2(t).$$

LEMMA 3: *Under Assumptions A-B and H_0 , $\text{plim}_{n \rightarrow \infty} \sup_{t \in T} |\hat{W}(t) - z_n(t)|^2 = 0$.*

LEMMA 4: *Under Assumptions A-B and H_0 , z_n is tight.*

It is easy to prove that for arbitrary t_1, \dots, t_q in T , $(z_n(t_1), \dots, z_n(t_q))'$

is asymptotically distributed as $(z(t_1), \dots, z(t_q))'$. Together with Lemma 4 this implies that z_n converges weakly to z . Cf. Billingsley (1968, p.47). The argument in Billingsley (1968) actually concerns the case $T = [0,1]$, but all the relevant results carry over to hypercubes. Since $(\cdot)^2$ and $\sup_{t \in T}(\cdot)$ are continuous mappings from $C(T)$ into $C(T)$, the conclusion of the theorem now follows from Lemma 3 and Theorem 5.1 of Billingsley (1968). Q.E.D.

Note that the covariance function $\Gamma(t_1, t_2)$ depends on the distribution of (y_j, x_j) and on the model, and so does the distribution of $\sup_{t \in T} z(t)^2$. Thus this null distribution has to be calculated each time we conduct this test on a different data set or for a different model. Another problem is how to calculate the null distribution involved. Therefore we propose the following alternative for Theorem 3:

THEOREM 4: Let Assumptions A-B hold. Choose independently of the data generating process real numbers $\gamma > 0$, $\rho \in (0,1)$, and a point $t_0 \in T$. Let $\hat{t} = \operatorname{argmax}_{t \in T} \hat{W}(t)$ and let

$$\tilde{t} = t_0 \text{ if } \hat{W}(\hat{t}) - \hat{W}(t_0) \leq \gamma n^\rho; \tilde{t} = \hat{t} \text{ if } \hat{W}(\hat{t}) - \hat{W}(t_0) > \gamma n^\rho.$$

Then under H_0 , $\hat{W}(\tilde{t}) \rightarrow \chi_1^2$ in distribution, whereas under H_1 , $\hat{W}(\tilde{t})/n \rightarrow \sup_{t \in T} \eta(t)$ a.s.

PROOF: The result under H_1 follows easily from Theorem 3. Assume now that H_0 is true. Theorem 3 implies that $\hat{W}(\hat{t}) - \hat{W}(t_0)$ is stochastically bounded, hence for every $\gamma > 0$, $\rho > 0$, $P[\hat{W}(\hat{t}) - \hat{W}(t_0) > \gamma n^\rho] \rightarrow 0$ and consequently, $\lim_{n \rightarrow \infty} P[\tilde{t} = t_0] = 1$. Since $\hat{W}(t_0) \rightarrow \chi_1^2$ in distribution, conditionally on t_0 , and t_0 is independent of the data generating process, the result for H_0 follows. Q.E.D.

In practice it may be quite laborious to determine \hat{t} . The following quick-and-easy procedure may serve as an alternative, not only for Theorem 4 but also for Theorem 3. The proof of this result is similar to the proof of Theorems 3 and 4.

THEOREM 5: Choose a sequence of positive integers K_n converging to infinity with n , and choose a sequence (t_i) such that $\{t_1, t_2, t_3, \dots\}$ is dense in T . Replace \hat{t} in Theorems 3 and 4 by $\hat{t} = \operatorname{argmax}_{t \in \{t_1, \dots, t_{K_n}\}} \hat{W}(t)$. Then Theorems 3 and 4 carry over.

There are many different ways to choose t_1, t_2, \dots , given the compact set T . For example, let $\{t_1, t_2, \dots\}$ be the set of rational-valued vectors in T . Also, one may choose the t_i 's ($i=0, 1, 2, \dots$) randomly from a continuous distribution with density having support T . In the latter case it should be stressed that the random search procedure involved differs fundamentally from the randomization procedures in Bierens (1987, 1988) and Bierens and Hartog (1988), as Theorem 5 holds true if we condition on the sequence (t_i) . In contrast, the aforementioned tests lose their consistency if one conditions on the random test parameters.

The choice of Φ

From an asymptotic point of view the choice of the bounded one-to-one mapping Φ is no issue. However, in Bierens (1982, 1984, 1987) we have advocated letting Φ depend on the scale of x_j . For if we choose Φ as in (5) and if the components x_{ij} of x_j take only large positive values then $\tan^{-1}(x_{ij}) \approx \frac{1}{2}\pi$ for $i=1, \dots, k$, hence $\exp[t'\Phi(x_j)] \approx \prod_{i=1}^k \exp(\frac{1}{2}\pi t_i)$. Clearly this will destroy the power of the test. A cure for this problem is to standardize the x_{ij} 's before taking the transformation Φ . Thus, replace the weighting function $\exp(t'\Phi(x_j))$ by

$$(19) \quad w(x_j, t) = \prod_{i=1}^k \exp[t_i \varphi((x_{ij} - \bar{x}_i)/s_i)],$$

where \bar{x}_i and s_i are the sample mean and the sample standard error of the x_{ij} , respectively, and φ is a bounded one-to-one mapping from \mathbb{R} into \mathbb{R} . If we choose for φ a continuously differentiable function with uniformly bounded first derivative, then it can be verified from the proofs of Theorems 3 and 4 that the resulting test has the asymptotic properties described in these theorems.

5. MONTE CARLO RESULTS

Next we show by a limited Monte Carlo simulation how the test in Theorem 4 in the form described in Theorem 5 performs in finite samples. To limit computation cost, we have chosen the integer K_n in Theorem 5 relatively small, namely $K_n = [n/10]-1$, and only 500 replications were used.

The data generating processes (DGP) we distinguish are the following. Let z_j , v_{1j} , v_{2j} and e_j be independent random drawings from the standard normal distribution, and let the regressors be $x_{1j}=z_j+v_{1j}$, $x_{2j}=z_j+v_{2j}$. The dependent variable is generated according to $y_j=1+x_{1j}+x_{2j}+u_j$, where either

$$\text{DGP 1: } u_j = v_{1j}v_{2j} + e_j, \text{ or } \text{DGP 2: } u_j = (\sqrt{2})e_j.$$

In both cases we fit a linear regression model:

$$H_0: y_j = \theta_0 + \theta_1 x_{1j} + \theta_2 x_{2j} + u_j, \quad E[u_j | x_{1j}, x_{2j}] = 0 \text{ a.s.}$$

Clearly, H_0 is false for DGP 1 and H_0 is true for DGP 2. Note that in both cases $E[u_j^2] = 2$, the theoretical coefficient of determination equals $2/3$ and $\text{plim}_{n \rightarrow \infty} \hat{\theta} = (1, 1, 1)'$.

In conducting the test we have chosen $T = [1, 5] \times [1, 5]$ and the t_i 's ($i=0, 1, \dots$) have been drawn randomly from the uniform distribution on T . We used the weighting function (19) with $\varphi(x) = \tan^{-1}(x/2)$. The Monte Carlo simulations have been conducted for sample sizes 50, 100, 200, 400 and 800 and four sets of values of the penalty parameters γ and ρ . The results of 500 replications of the DGP's involved are presented in Table 1.

< Insert Table 1 about here >

We see from Table 1 that the test seems rather insensitive to variations in the penalty γn^ρ , except in the last case where the penalty becomes too low, which affects the actual size of the test too much. In the first three cases the penalty is hardly effective under H_1 for $n \leq 400$ and in the first two cases also for $n=800$. This indicates that the function $\eta(t)$ is relatively flat. The finite sample power for $n = 400$ at the 5% significance level is still not equal to 1, although (as expected) the power increases with n . However, for $n=800$ the power gets close to 1. The variation in the actual size of the test (apart from the last case) may be due to the limited amount of Monte Carlo simulations.

< Insert Figure 1 about here >

In order to understand better these results, we subsequently have calculated the function $\eta(t)$ for DGP1. This was done partially analytically, partially by simulation. Figure 1 shows the shape of $\eta(t)$ for $-5 \leq t_1 \leq 5$, $-5 \leq t_2 \leq 5$, looking from above at an angle of about 60 degrees. Note that $\eta(t)$ is symmetric about the diagonal $(-5,-5)-(5,5)$. Moreover, $\eta(t)$ equals zero on the two axes $t_1=0$ and $t_2=0$, as can easily be proved. Thus, these two axes form the set S in Theorem 2. Furthermore, the set S_* in Lemma 2 consists of the origin $(0,0)$ only. The height of the hills in the quadrants $[-5,0] \times [0,5]$ and $[0,5] \times [-5,0]$ is about .02. The flat hill in the quadrant $[0,5] \times [0,5]$ is just located in the area T we have chosen for the Monte Carlo simulations. Its height is about .004, i.e., only 20 % of the maximum height of the surface. The relative flatness of this hill is the very reason why in the first three cases in Table 1 the penalty remained effective under H_1 , for (loosely speaking) t_0 will prevail as long as $\sup_{t \in T} \eta(t) - \eta(t_0) \leq \gamma n^{\rho-1}$. Apparently we have accidentally selected T in one of the worst areas. However, T should be chosen independently of the data generating process, i.e., it is not allowed to determine the best set T by looking at the plot of $\hat{W}(t)$.

In cross-section analysis we often work with much larger samples than in Table 1. For such samples the test will likely work according to the prediction of asymptotic theory. Presumably the small sample power of our test will be inferior to the small sample power of a test designed to test consistently H_0 against a specific alternative model, as it likely trades away small sample power against any one alternative for consistency against all alternatives. Also, the performance of more general consistent CM tests may be better than the present one, as the additional weights may enhance the finite sample power.

6. WHAT INFORMATION DOES THE TEST REVEAL IF H_0 IS REJECTED ?

If H_0 is false, then the function $\eta(t)$ contains information about the true model. But what kind of information ? To answer this query, denote for $i=1,2,\dots$,

$$\hat{u}_j = y_j - f(x_j, \hat{\theta}), \quad \hat{w}_{1j} = \exp(t_1' \Phi(x_j)) - b(t_1)' \hat{A}^{-1} (\partial / \partial \theta') f(x_j, \hat{\theta}).$$

Then the test statistic $\hat{W}(t)$ in Theorems 3 and 5 is just the most significant squared t-value of the parameters λ_i in the linear regressions

$$\hat{u}_j = \lambda_i \hat{w}_{ij} + \hat{v}_{ij}, \quad i=1, \dots, K_n,$$

provided the t-values involved are calculated according to the approach of White (1980). Similarly, denoting

$$w_{ij} = \exp(t_i' \Phi(x_j)) - b(t_i)' A^{-1} (\partial / \partial \theta') f(x_j, \theta_0),$$

the test involved selects asymptotically the alternative model

$$(20) \quad y_j = f(x_j, \theta_0) = \lambda_i w_{ij} + v_{ij}, \quad i=1, 2, \dots,$$

for which the probability limit $\eta(t_i)$ of the squared t-value of λ_i , divided by n , is maximal. In other words, if H_1 is true the test augments the model with the most significant additional regressor w_{ij} . The test in Theorems 4 and 5 does the same, except that now $i=0, 1, 2, \dots$ and that choosing w_{ij} for $i > 0$ is penalized, by which under H_0 the test statistic is asymptotically equivalent to the squared t-value of λ_0 . Note that under H_1 the augmented model (20) is not necessarily the true model. It is only closer to the true model, in terms of quadratic loss, than the original model.

The above argument suggests a further elaboration of the test by regressing \hat{u}_j on $\hat{w}_{1j}, \dots, \hat{w}_{K_n, j}$, where K_n is determined by some selection criterion for model dimension like the Akaike (1974) and Schwarz (1978) criteria. Under H_0 we may then expect that K_n will converge to 1, so that similarly to Theorems 4 and 5 the null distribution of the Wald test of the joint significance of the K_n parameters involved is asymptotically χ_1^2 . However, this further elaboration is beyond the scope of the present paper.

Department of Econometrics, Free University, 1081HV Amsterdam, The Netherlands.

First draft: January 1987.

Revision: September 1989

APPENDIX A: Maintained Assumptions

The following assumptions are standard for consistency and asymptotic normality of nonlinear least squares estimators. Cf. Jennrich (1969) and White (1980, 1981, 1982).

(A.1) Let $\{(y_1, x_1), \dots, (y_n, x_n)\}$ be a sample from a probability distribution $F(y, x)$ on $R \times R^k$. Moreover, $E y_j^2 < \infty$.

(A.2) The parameter space Θ is a compact and convex subset of R^m and $f(x, \theta)$ is for each $\theta \in \Theta$ a Borel measurable real function on R^k and for each k -vector x a twice continuously differentiable real function on Θ . Moreover, $E[\sup_{\theta \in \Theta} f(x_j, \theta)^2] < \infty$ and for $i_1, i_2 = 1, \dots, m$,

$$E[\sup_{\theta \in \Theta} |((\partial/\partial \theta_{i_1})f(x_1, \theta))((\partial/\partial \theta_{i_2})f(x_1, \theta))|] < \infty,$$

$$E[\sup_{\theta \in \Theta} |(y_1 - f(x_1, \theta))^2 ((\partial/\partial \theta_{i_1})f(x_1, \theta))((\partial/\partial \theta_{i_2})f(x_1, \theta))|] < \infty,$$

$$E[\sup_{\theta \in \Theta} |(y_1 - f(x_1, \theta))(\partial/\partial \theta_{i_1})(\partial/\partial \theta_{i_2})f(x_1, \theta)|] < \infty.$$

(A.3) $E[(y_1 - f(x_1, \theta))^2]$ takes a unique minimum on Θ at θ_0 . Under H_0 the parameter vector θ_0 is an interior point of Θ .

(A.4) The matrix A defined in (11) is non-singular.

APPENDIX B: Proofs of the Lemmas

PROOF OF LEMMA 1: First, let $k=1$. According to Theorem 2 of Bierens (1982) there exists a nonnegative integer m such that $E[v \cdot x^m] \neq 0$, hence

$$(d/dt)^m E[v \cdot e^{t \cdot x}] = \sum_{j=m}^{\infty} t^{j-m} E[v \cdot x^j] / (j-m)! \rightarrow E[v \cdot x^m] \neq 0 \text{ as } t \rightarrow 0.$$

This implies that $E[v \cdot \exp(tx)] \neq 0$ in a neighborhood of zero. Now let t_0 be such that $E[v \cdot \exp(t_0 x)] = 0$. Since $P\{E[v \cdot \exp(t_0 x) | x] = 0\} = P\{E[v | x] = 0\} < 1$, it follows from the above argument, with v replaced by $v \cdot \exp(t_0 x)$, that $E[v \cdot \exp(t_0 \cdot x + t \cdot x)] \neq 0$ in a neighborhood of $t = 0$, hence $E[v \cdot \exp(t \cdot x)] \neq 0$ in a neighborhood of $t = t_0$. This implies $\inf_{t \in S, t \neq t_0} |t - t_0| > 0$ if $t_0 \in S$ and hence that S is countable. Since a countable set has Lebesgue measure zero, the lemma follows for the case $k = 1$.

Next, consider the general case $k \geq 1$. Let $x = (x_1, x_2, \dots, x_k)'$, $t = (t_1, t_2, \dots, t_k)'$. Again it follows from Theorem 2 of Bierens (1982) that there exist nonnegative integers m_1, \dots, m_k such that $E[v \cdot x_1^{m_1} x_2^{m_2} \dots x_k^{m_k}] \neq 0$ and similarly to the case $k = 1$ this implies that there exists a t_* close to the origin of R^k such that $E[v \cdot \exp(t_*' x)] \neq 0$. Let $v_* = v \cdot \exp(t_*' x)$. Since $E v_* \neq 0$ we have $P\{E[v_* | x_1, \dots, x_k] = 0\} < 1$ for

$\ell = 1, \dots, k$. Now suppose that for some $\ell \geq 1$ the set

$$S_\ell^* = \{(t_1, \dots, t_\ell)' \in \mathbb{R}^\ell : E[v_* \exp(t_1 x_1 + \dots + t_\ell x_\ell)] = 0\}$$

has Lebesgue measure zero. Then by the argument for the case $k = 1$ the set

$$S_{\ell+1}^{**}(t_1, \dots, t_\ell) = \{t \in \mathbb{R} : E[(v_* \exp(t_1 x_1 + \dots + t_\ell x_\ell)) \exp(t x_{\ell+1})] = 0\}$$

is countable if $(t_1, \dots, t_\ell) \notin S_\ell^*$, whereas $S_{\ell+1}^{**}(t_1, \dots, t_\ell) = \mathbb{R}$ if $(t_1, \dots, t_\ell) \in S_\ell^*$. Since now $S_{\ell+1}^{**}$ is the union of two sets with Lebesgue measure zero, namely the sets $\{(t_1, \dots, t_{\ell+1}) : (t_1, \dots, t_\ell) \notin S_\ell^*, t_{\ell+1} \in S_{\ell+1}^{**}(t_1, \dots, t_\ell)\}$ and $S_\ell^* \times \mathbb{R}$, it has Lebesgue measure zero itself. By induction it follows that S_k^* has Lebesgue measure zero. Replacing t by $t - t_*$ the lemma follows. Q.E.D.

PROOF OF LEMMA 2: Suppose that S_* has positive Lebesgue measure. For any $t \in S_*$ we have $\sigma^2(x_1)[\exp(t' \Phi(x_1)) - b(t)' A^{-1}(\partial/\partial \theta') f(x_1, \theta_0)]^2 = 0$ a.s., hence by the first part of Assumption B,

$$\exp(t' \Phi(x_1)) - b(t)' A^{-1}(\partial/\partial \theta') f(x_1, \theta_0) = 0 \text{ a.s.}$$

and consequently, using (10),

$$\begin{aligned} E[\mu(x_1) \exp(t' \Phi(x_1))] &= E[\mu(x_1) (\partial/\partial \theta) f(x_1, \theta_0)] A^{-1} b(t) \\ &= E[\lambda' (\partial/\partial \theta') f(x_1, \theta_0) \exp(t' \Phi(x_1))], \end{aligned}$$

where $\lambda = A^{-1} E[\mu(x_1) (\partial/\partial \theta') f(x_1, \theta_0)]$. Since this result holds for all t in a set with positive Lebesgue measure, it follows from Lemma 1 that $\mu(x_1) = \lambda' (\partial/\partial \theta') f(x_1, \theta_0)$ a.s., hence $E[\kappa \kappa']$ is singular. This contradicts with the second part of Assumption B. Q.E.D.

PROOF OF LEMMA 3: Let $\hat{M}(\theta, t) = (1/n) \sum_{j=1}^n (y_j - f(x_j, \theta)) \exp[t' \Phi(x_j)]$. By the mean value theorem we have

$$(B.1) \quad \sqrt{n} \hat{M}(\hat{\theta}, t) - \sqrt{n} \hat{M}(\theta_0, t) = [(\partial/\partial \theta') \hat{M}(\bar{\theta}^{(1)}(t), t)]' \sqrt{n}(\hat{\theta} - \theta_0),$$

where $\tilde{\theta}^{(1)}(t)$ is a mean value satisfying $|\tilde{\theta}^{(1)}(t) - \theta_0| \leq |\hat{\theta} - \theta_0|$ a.s. Let $\hat{b}(\theta, t) = -(\partial/\partial\theta')\hat{M}(\theta, t)$. From Assumption A, the consistency of $\hat{\theta}$ and the uniform law of large numbers of Jennrich (1969) it easily follows

$$(B.2) \quad \text{plim}_{n \rightarrow \infty} \sup_{t \in T} |\hat{b}(\tilde{\theta}^{(1)}(t), t) - E \hat{b}(\theta_0, t)| = 0.$$

Observe that $E \hat{b}(\theta_0, t) = b(t)$, hence it follows from (B.1) and (B.2),

$$(B.3) \quad \text{plim}_{n \rightarrow \infty} \sup_{t \in T} |\sqrt{n}\hat{M}(\hat{\theta}, t) - \sqrt{n}\hat{M}(\theta_0, t) + b(t)'\sqrt{n}(\hat{\theta} - \theta_0)| = 0.$$

Furthermore, from standard nonlinear least squares theory it follows

$$(B.4) \quad \text{plim}_{n \rightarrow \infty} |\sqrt{n}(\hat{\theta} - \theta_0) - A^{-1}(1/\sqrt{n})\sum_{j=1}^n u_j (\partial/\partial\theta')f(x_j, \theta_0)| = 0.$$

where $u_j = y_j - f(x_j, \theta_0)$. Substituting (B.4) in (B.3) yields

$$(B.5) \quad \text{plim}_{n \rightarrow \infty} \sup_{t \in T} |\sqrt{n}\hat{M}(\hat{\theta}, t) - z_n(t)/s^2(t)| = 0$$

Observe that by the uniform law of Jennrich (1969) and Assumption A,

$$(B.6) \quad \text{plim}_{n \rightarrow \infty} \sup_{t \in T} |\hat{s}^2(t) - s^2(t)| = 0.$$

Since by assumption, $\inf_{t \in T} s^2(t) > 0$, the lemma follows from (B.5) and (B.6). Q.E.D.

PROOF OF LEMMA 4: According to Theorem 8.2 of Billingsley (1968) it suffices to prove:

$$(B.7) \quad \text{For each } \delta > 0 \text{ and an arbitrary } t_0 \in T \text{ there exists an } \epsilon \text{ such that} \\ \sup_n P(|z_n(t_0)| > \epsilon) \leq \delta;$$

$$(B.8) \quad \text{For each } \delta > 0 \text{ and } \epsilon > 0 \text{ there exists an } \xi > 0 \text{ such that} \\ \sup_n P[\sup_{|t_1 - t_2| < \xi} |z_n(t_1) - z_n(t_2)| \geq \epsilon] \leq \delta.$$

Condition (B.7) follows from the fact that $z_n(t_0) \rightarrow N(0,1)$ in distribution. For proving condition (B.8), assume for the moment that t is scalar. Then (B.8) follows from

$$\begin{aligned}
 (B.9) \quad & E(\sup_{|t_1 - t_2| < \xi} |(1/\sqrt{n}) \sum_{j=1}^n u_j [\exp(t_1 \Phi(x_j)) - \exp(t_2 \Phi(x_j))]|) \\
 & \leq \sum_{i=1}^{\infty} E |(1/\sqrt{n}) \sum_{j=1}^n u_j \Phi(x_j)^i| \sup_{|t_1 - t_2| < \xi} |t_1 - t_2|^i / i! \\
 & \leq \sqrt{(E(u_1^2))} \sum_{i=1}^{\infty} [2\xi \cdot \sup_x |\Phi(x)|]^i / i! = \sqrt{(E(u_1^2))} (\exp[2\xi \cdot \sup_x |\Phi(x)|] - 1)
 \end{aligned}$$

and the fact that $b(t)$ is continuous. The proof for the case that x_j and t are vectors goes along the same line, using the multinomial expansion of $[t' \Phi(x_j)]^i$ in (B.9). Q.E.D.

REFERENCES:

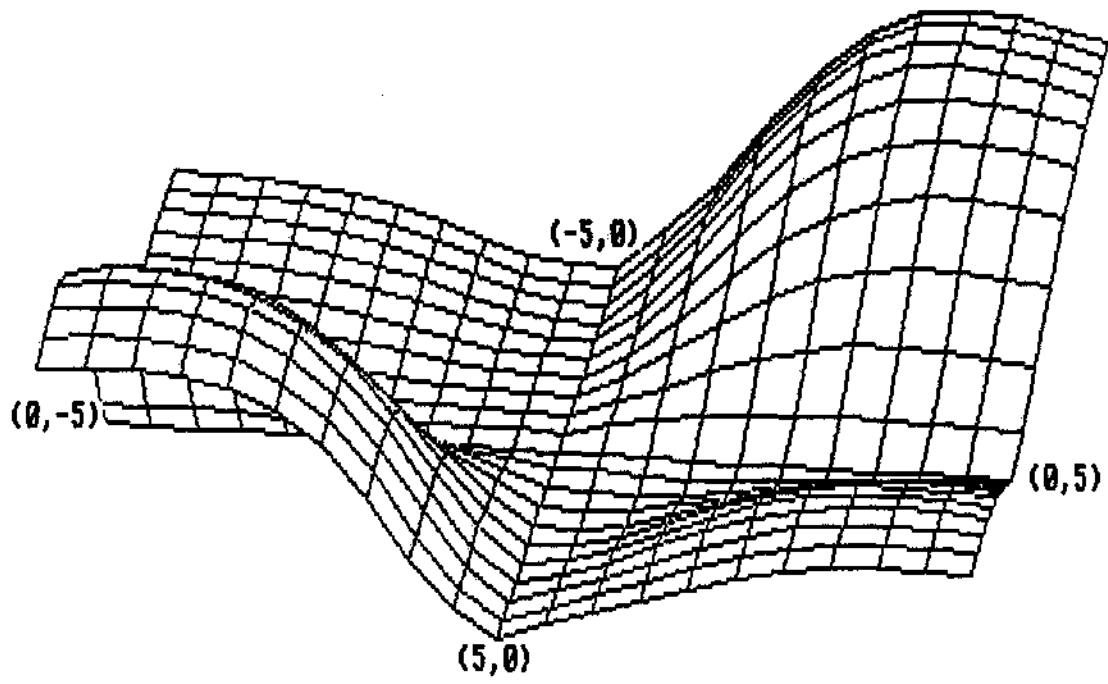
- AKAIKE, H. (1974): "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control* 19, 716-723
- BIERENS, H.J. (1982): "Consistent Model Specification Tests", *Journal of Econometrics* 20, 105-134
- BIERENS, H.J. (1984): "Model Specification Testing of Time Series Regressions", *Journal of Econometrics* 26, 323-353
- BIERENS, H.J. (1987): "ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands", *Journal of Econometrics* 35, 161-190
- BIERENS, H.J. (1988): "ARMA Memory Index Modeling of Economic Time Series (with discussion)", *Econometric Theory* 4, 35-59
- BIERENS, H.J. and J. HARTOG (1988): "Nonlinear Regression with Discrete Explanatory Variables, with an Application to the Earnings Function", *Journal of Econometrics* 38, 269-299
- BILLINGLEY, P. (1968): *Convergence of Probability Measures*, New York: Wiley
- EL BADAWI, I., A.R. GALLANT and G. SOUZA (1983): "An Elasticity Can Be Estimated Consistently without A Priori Knowledge of Functional Form", *Econometrica* 51, 1731-1735
- FRIEDMAN, J.H. and W. STUETZLE (1981): "Projection Pursuit Regression", *Journal of the American Statistical Association* 76, 817-823
- GALLANT, A.R. (1981): "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form", *Journal of Econometrics* 15, 211-245

- GALLANT, A.R. (1982): "Unbiased Determination of Production Technologies", *Journal of Econometrics* 20, 285-323
- GALLANT, A.R. (1984): "The Fourier Flexible Form", *American Journal of Agricultural Economics* 66, 204-208.
- HAUSMAN, J.A. (1978): "Specification Testing in Econometrics", *Econometrica* 46, 1251-1271
- HOLLY, A. (1982): "A Remark on Hausman's Specification Test", *Econometrica* 50, 749-760
- HUBER, P. (1985): "Projection Pursuit (with discussion)", *Annals of Statistics* 13, 435-475
- JENNRICH, R.I. (1969): "Asymptotic Properties of Nonlinear Least Squares Estimators", *Annals of Mathematical Statistics* 40, 633-643
- RAMSEY, J.B. (1969), "Tests for Specification Errors in Classical Linear Least-squares Regression Analysis", *Journal of the Royal Statistical Society, Series B* 31, 350-371
- RAMSEY, J.B. (1970), "Models, Specification Error, and Inference: A Discussion of Some Problems in Econometric Methodology", *Bulletin of the Oxford Institute of Economics and Statistics* 32, 301-318
- ROYDEN, H.L. (1968): *Real Analysis*. London: Macmillan
- SCHWARZ, G. (1978): "Estimating the Dimension of a Model", *Annals of Statistics* 6, 461-464
- NEWKEY, W.K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica* 53, 1047-1070
- RUUD, P.A. (1984): "Tests of Specification in Econometrics", *Econometric Reviews* 3, 211-242
- WHITE, H. (1980): "Nonlinear Regression on Cross-Section Data", *Econometrica* 48, 721-746
- WHITE, H. (1981): "Consequences and Detection of Misspecified Nonlinear Regression Models", *Journal of the American Statistical Association* 76, 419-433
- WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models", *Econometrica* 50, 1-25

Table 1: Monte Carlo Results (500 Replications)

Test parameters γ, ρ	n	% Rejections of H_0				Mean (S.E.) of the test statistic				Number of times $\tilde{t}=t_0$	
		5% Signif.		10% Signif.							
		DGP 1	DGP 2	DGP 1	DGP 2	DGP 1		DGP 2		DGP 1	DGP 2
$\gamma=1.0$ $\rho=0.5$	50	8.4	1.4	21.4	6.0	1.640	(1.58)	0.977	(0.98)	500	500
	100	18.2	2.4	30.6	6.6	2.154	(1.90)	0.933	(1.11)	500	500
	200	29.0	3.8	49.0	9.0	3.065	(2.34)	0.992	(1.19)	500	500
	400	64.2	3.8	76.2	7.2	5.582	(3.84)	0.932	(1.25)	500	500
	800	89.6	3.8	95.4	8.8	9.807	(5.46)	0.998	(1.22)	499	500
$\gamma=0.5$ $\rho=0.5$	50	8.4	2.6	19.0	6.8	1.597	(1.52)	0.960	(1.07)	495	499
	100	16.6	4.0	29.8	9.4	2.186	(1.89)	1.072	(1.22)	498	498
	200	31.0	3.8	47.0	8.2	3.256	(2.85)	1.002	(1.22)	489	500
	400	63.0	3.4	79.0	9.8	5.860	(4.05)	0.996	(1.24)	481	500
	800	88.6	4.4	94.6	10.2	12.13	(7.97)	0.984	(1.22)	429	500
$\gamma=0.25$ $\rho=0.5$	50	12.8	4.0	24.4	12.4	1.849	(1.69)	1.167	(1.22)	462	475
	100	25.2	6.0	35.4	10.0	2.537	(2.45)	1.149	(1.40)	452	487
	200	46.2	5.6	57.6	12.0	4.206	(3.43)	1.155	(1.41)	412	494
	400	75.6	6.0	83.4	10.8	8.288	(5.30)	1.080	(1.53)	330	495
	800	94.8	4.4	97.6	10.6	17.22	(8.09)	0.981	(1.29)	185	500
$\gamma=0.25$ $\rho=0.25$	50	12.8	5.2	27.4	10.2	2.028	(1.60)	1.225	(1.27)	343	389
	100	26.2	6.0	45.8	13.4	2.861	(2.20)	1.300	(1.31)	281	328
	200	59.0	9.0	72.6	20.0	5.105	(3.31)	1.580	(1.65)	178	367
	400	91.4	10.0	96.4	17.4	9.675	(4.66)	1.465	(1.77)	79	379
	800	100	8.2	100	16.2	18.71	(6.29)	1.265	(1.51)	20	407

Figure 1



Footnotes:

- 1) The helpful comments of Lars Peter Hansen and four referees, leading to substantial improvements over previous versions of this paper, are gratefully acknowledged.
- 2) Throughout this paper we denote by $|\cdot|$ the absolute value if the argument is a scalar and the Euclidean norm if the argument is a vector.
- 3) Following Bierens (1982) one may also take $\int_1^{\hat{W}} \hat{W}(t) dt$ as test statistic. Theorem 3 implies that under H_0 this test statistic converges in distribution to $\int_1 z(t)^2 dt$. Note that the convergence in distribution of the corresponding test statistic in Bierens (1982) was not proved.

